



# On the scaling patterns of species spatial distribution and association

Cang Hui \*

Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

## ARTICLE INFO

### Article history:

Received 6 May 2009

Received in revised form

12 August 2009

Accepted 14 August 2009

Available online 21 August 2009

### Keywords:

Spatial scale

Autocorrelation

Co-occurrence

Join-count statistics

Pair approximation

Bayesian model

Aggregation

Spatial dissociation

Beta diversity

## ABSTRACT

Understanding how species distribution (occupancy and spatial autocorrelation) and association (that is, multi-species co-distribution) change across scales is fundamental to unlocking the pattern formation in population ecology and macroecology. Based on the Bayesian rule and join-count statistics, I present here a mathematical model that can demonstrate the effect of spatial scale on the observation of species distribution and association. Results showed that the intensity of spatial autocorrelation and species association declines when the grain in the spatial analysis increases, although the category of species distribution (aggregated or segregated) and association (positive or negative) remains the same. Random distribution and species independence were proved to be scale-free. Regardless of the possible patterns of species distribution and association, species tend to be randomly distributed and independent from each other when scaling-up (an increasing grain), reflecting a percolation process. This model, thus, grasps the statistical essence of species scaling pattern and presents a step forward for unveiling mechanisms behind species distributional and macroecological patterns.

© 2009 Elsevier Ltd. All rights reserved.

## 1. Introduction

Unveiling the relationship between patterns and scales is of fundamental importance in population ecology and macroecology (Gaston and Blackburn, 2000; Levin, 1992; Wu, 2004). Predicting species distribution, in terms of occupancy and spatial autocorrelation, and association (co-occurring and overlapping of species distribution) across scales contributes much to extracting ecological mechanisms from sampling and statistical artefacts (e.g., Hui and McGeoch, 2007a; McGeoch and Gaston, 2002), and thus further advances our knowledge of pattern formation in ecological communities (Storch et al., 2008) and its applications in biological conservation (Wilson et al., 2004). However, despite their paramount importance, species distribution and association remains difficult to dissect and estimate across scales (Bell, 2005; He and Gaston, 2003).

The distribution of a species can be depicted as its range size (occupancy) and spatial structure (e.g. spatial autocorrelation). Species occupancy, measured as the number of presences in samples, localities or lattices, forms the basis for assessing species conservation status and viability (Gaston and Fuller, 2009; McGeoch and Gaston, 2002). Unlocking macroecological patterns, such as species occupancy–abundance relationship (Holt et al.,

2002; Hui and McGeoch, 2007b) and species occupancy frequency distribution (Hui and McGeoch, 2007a,c; McGeoch and Gaston, 2002), is largely dependent on our understanding of how the spatial scale affects species occupancy in ecological surveys. The other crucial variable, as mentioned above, for depicting species distribution is the non-randomness, or aggregation, of its spatial structure. Species are normally not randomly distributed in space, but aggregated. This aggregation can arise both from environmental heterogeneity (Fahrig and Nuttle, 2005) and the non-linearity of biological processes (Pacala and Levin, 1997). Wiens (2000) distinguished four forms of non-randomness, namely spatial, patterned, compositional and locational variances. Of special importance to population ecology is the pattern variance, which measures the spatial autocorrelation of species distribution and has been shown to be highly scale-dependent (Hurlbert, 1990; Jelinski and Wu, 1996; Openshaw, 1984; Perry et al., 2002). Describing this pervasive, non-random spatial aggregation remains a fundamental aspect of ecological research (Fortin and Dale, 2005). Without a clear understanding of the spatial scaling properties of aggregation, many of the most interesting and critical problems in ecology cannot be fully addressed. This is because the scaling pattern of species aggregation is intricately entangled with patterns in species ranges, abundance and diversity (Gaston and Blackburn, 2000; Gaston et al., 2006; Ives, 1991; Shorrocks and Sevenster, 1995; Wertheim et al., 2000).

Species are not alone but associated positively or negatively (also known as spatial dissociation) with each other, as a result of,

\* Tel.: +27 21 808 3413; fax: +27 21 808 2995.

E-mail address: [chui@sun.ac.za](mailto:chui@sun.ac.za)

for instance, competition or facilitation in communities. Species association belongs to [Wiens's \(2000\)](#) compositional variance and measures the degree of co-occurrence and co-distribution of two species in samples ([Bell, 2005](#); [Gotelli and Graves, 1996](#)). The scaling pattern of species association is closely tied with patterns of beta diversity and species-area relationship (e.g. [Harte et al., 2005](#); [Hui and McGeoch, 2008](#)) and plays a vital role in identifying species assembly patterns and interactions from predictions solely be chance or null models ([Connor and Simberloff, 1979](#); [Diamond, 1975](#); [Gotelli and Graves, 1996](#)). The scaling pattern of occupancy, spatial autocorrelation and species association, together, not only allows us to provide a comprehensive description of species distribution in ecological communities, but also can evaluate species role and function in forming ecosystems.

Here, I present a mathematical model demonstrating the effect of spatial scale on these three variables of species distribution, i.e. a model of the scaling patterns of species occupancy, autocorrelation and association. This model is based on the Bayesian rule ([Hui et al., 2006](#)) as well as the paralleling definitions of the pair approximation in pattern recognition ([Iwasa et al., 1991](#); [Matsuda et al., 1992](#); [Sato and Iwasa, 2000](#)) and the join-count statistics in spatial analysis ([Fortin et al., 2002](#); [Fortin and Dale, 2005](#)). Although both pattern recognition and spatial analysis aim to achieve a common goal (that is, to quantify the non-random patterns of species distributions), research from these two different fields barely intersects. For instance, pair approximation is initiated from physics and often used as a method for describing the spatial patterns from lattice models in theoretical ecology ([Ellner, 2001](#); [Sato and Iwasa, 2000](#)), whereas join-count statistics is a method for quantifying species spatial autocorrelation with data collected from real ecological surveys ([Fortin et al., 2002](#)). By presenting this model, I thus emphasize the importance and potential of bringing these similar techniques from different fields together in understanding ecological complexities ([Hui, 2008](#)).

## 2. Model

### 2.1. Join-count statistics

Species distribution can be defined as the geographic range of its populations or the spatial arrangement of individuals in a local population ([Molles, 1999](#)). A standard format of reporting a species' distribution is the presence-absence map as used in most species atlas ([Gaston and Fuller, 2009](#)). This binary-format map is, mathematically, equal to a binary matrix with the element being either 0 (absence) or 1 (presence). This kind of matrices is normally obtained in ecology from two sources: cellular automata (or lattice models) in theoretical ecology and grid sampling from ecological surveys. The simplest way of describing the spatial structure of a binary matrix is the join-count statistics (see below; [Fortin et al., 2002](#)), which is conceptually and mathematically similar to the pair approximation (or moment approximation). [Sato and Iwasa \(2000\)](#) have presented an excellent introduction of using pair approximation in analysing spatial patterns from lattice models.

In a binary matrix, the global density of a population is defined as the probability that a randomly chosen element (called a cell hereafter) is occupied by the species. It has the same meaning as species occupancy,  $p_+(a)$ , where  $a$  indicates the grain, i.e. the size of the cell. The probability of absence is thus  $p_0(a) = 1 - p_+(a)$ . The local density of a species  $q_{+/+}(a)$  is the conditional probability that a randomly chosen adjacent cell of an occupied cell is also occupied, which is the simplest measure of spatial autocorrelation (at the scale of  $a$ ) among the local indicators of spatial association ([Anselin, 1995](#); [Fortin and Dale, 2005](#)). There are another three

conditional probabilities regarding the state of a focal cell and its neighbouring cells: an occupied cell with an empty neighbour,  $q_{0/+}(a)$ ; an empty cell with an occupied neighbour,  $q_{+/0}(a)$ ; an empty cell with an empty neighbour,  $q_{0/0}(a)$ . These four conditional probabilities form the join-count statistics for a single species ([Fortin et al., 2002](#)). However, we only need two variables,  $p_+(a)$  and  $q_{+/+}(a)$ , to express all other probabilities:

$$q_{0/+}(a) = 1 - q_{+/+}(a)$$

$$q_{+/0}(a) = (1 - q_{+/+}(a))p_+(a)/(1 - p_+(a))$$

$$q_{0/0}(a) = (1 - 2p_+(a) + q_{+/+}(a)p_+(a))/(1 - p_+(a))$$

Additional inequalities control the feasible range of these two variables:  $0 \leq p_+(a) \leq 1$  and  $2 - 1/p_+(a) \leq q_{+/+}(a) \leq 1$  ([Hui and Li, 2004](#)). A spatial aggregation can be defined as  $q_{+/+}(a) > p_+(a)$ , indicating a positive first-order spatial correlation between two local populations inhabiting adjacent cells. The random distribution is defined as  $q_{+/+}(a) = p_+(a)$  and implies the independence of two local populations in adjacent cells. The spatial segregation can be depicted by  $q_{+/+}(a) < p_+(a)$ , i.e. a negative spatial correlation between two adjacent occupied cells.

### 2.2. Species association

The above join-count statistics can only provide descriptions for one species in a binary matrix. If we consider two species, there exist four scenarios for a randomly chosen cell: species A and B coexist,  $P_{A \cap B}(a)$  (called the joint occupancy); only species A occurs,  $P_{A \cap \bar{B}}(a)$ ; only species B occurs,  $P_{\bar{A} \cap B}(a)$ ; neither occurs,  $P_{\bar{A} \cap \bar{B}}(a)$ . Knowing any one of these four scenarios, we can calculate the probabilities of other three. For instance, knowing the joint occupancy  $P_{A \cap B}(a)$ , we have

$$P_{A \cap \bar{B}}(a) = P_A(a) - P_{A \cap B}(a),$$

$$P_{\bar{A} \cap B}(a) = P_B(a) - P_{A \cap B}(a),$$

$$P_{\bar{A} \cap \bar{B}}(a) = 1 - P_A(a) - P_B(a) + P_{A \cap B}(a),$$

where  $P_A(a)$  and  $P_B(a)$  are the occupancy  $p_+(a)$  of species A and B, respectively. Therefore, we can define a positive association between species A and B as  $P_{A \cap B}(a) > P_A(a) \times P_B(a)$ , indicating a higher joint occupancy than expected when species A and B are independent from each other ( $P_A(a) \times P_B(a)$ ). Consequently, a negative association can be defined as  $P_{A \cap B}(a) < P_A(a) \times P_B(a)$ , and the independence between these two species  $P_{A \cap B}(a) = P_A(a) \times P_B(a)$ . This definition of species association is also closely linked with the C-score test in examining species co-distribution ([Bell, 2005](#); [Gotelli and Graves, 1996](#); [Stone and Roberts, 1990](#)). In the following, I present a model governing the scaling patterns of the occupancy  $p_+(a)$ , local density  $q_{+/+}(a)$  and the joint occupancy  $P_{A \cap B}(a)$ .

### 2.3. A Bayesian model

Let us consider species distribution when combining four adjacent cells (with coordinates  $\{x,y\}$ ,  $\{x+1,y\}$ ,  $\{x,y+1\}$ ,  $\{x+1,y+1\}$ ) as one larger cell, i.e. increasing the grain from  $a$  to  $4a$ . A  $4a$ -size cell consists of four  $a$ -size sub-cells and can be either absence (empty in all four sub-cells) or presence (with only 1, 2, 3 or 4 sub-cells being occupied). It is convenient to calculate the probability of absence  $p_0(4a)$ , instead of the more complicated probability of presence  $p_+(4a) = p_1(4a) + p_2(4a) + p_3(4a) + p_4(4a)$ , where  $p_i(4a)$  is the probability of  $i$  number of sub-cells occupied by the species. To calculate the probability of absence in a  $4a$ -size cell, we let the probability of absence  $p_0(a)$  for a randomly chosen sub-cell  $\{x,y\}$  multiplied by twice the conditional probability of its adjacent cells ( $\{x+1,y\}$  and  $\{x,y+1\}$ ) also being empty  $q_{0/0}(a)$  and then multiplied by the probability that a chosen sub-cell  $\{x+1,y+1\}$  with

two absent adjacent sub-cells is also found absent  $b_0(a)$ , and yield,

$$p_0(4a) = p_0(a) \times q_{0/0}(a)^2 \times b_0(a). \quad (1)$$

The calculation of the correlation between two adjacent empty cells can follow the same procedure:

$$q_{0/0}(4a) = q_{0/0}(a)^2 \times b_0(a)^2. \quad (2)$$

In the above equations,  $b_0(a)$  is the only probability unknown. Here, it is estimated using the Bayesian rule:

$$b_0(a) = \frac{q_{0/0}(a)^2 \times p_0(a)}{q_{0/0}(a)^2 \times p_0(a) + q_{0/+}(a)^2 \times p_+(a)}. \quad (3)$$

Because  $p_+(4a) = 1 - p_0(4a)$  and  $q_{+/+}(4a) = 1 - (1 - q_{0/0}(4a) \times p_0(4a) / p_+(4a))$ , we have the following occupancy and local density when the grain increases from  $a$  to  $4a$ :

$$p_+(4a) = 1 - \frac{\nabla^4}{\Delta}, \quad (4)$$

$$q_{+/+}(4a) = \frac{\nabla^{10} - 2\nabla^4 \Delta^2 + \Delta^3}{\Delta^2(\Delta - \nabla^4)}, \quad (5)$$

where  $\nabla = p_0(a) - q_{0/+}(a)p_+(a)$  and  $\Delta = p_0(a)[1 - p_+(a)^2(2q_{+/(+)}(a) - 3) + p_+(a)(q_{+/(+)}(a)^2 - 3)]$ . Studies using *Drosophilidae* in a decaying fruit mesocosm (Hui et al., 2006) and using the atlas of southern African birds (Hui et al., 2009) to predict occupancy and local density across scales have verified the robustness and accuracy of the above formula for the scaling patterns of species occupancy and spatial autocorrelation.

The above model provided a discrete version of the scaling patterns of species occupancy and autocorrelation. A continuous version of the scaling patterns can also be induced. Analogous to Eqs. (1) and (2), we have the following probabilities when combining  $n \times n$  cells together in forming a larger cell:

$$p_0(n \times n) = p_0 \times q_{0/0}^{2(n-1)} \times b_0^{(n-1)^2}, \quad (6)$$

$$q_{0/0}(n \times n) = q_{0/0}^n \times b_0^{n(n-1)}, \quad (7)$$

where, for simplicity,  $p_0$  denotes the absence for a unit-size cell  $p_0(1)$ , and so are  $q_{0/0}$  and  $b_0$ . Since  $n \times n$  indicates the size of grain,  $a$ , we can have the following scaling patterns from Eqs. (6) and (7) by replacing  $n$  with  $a^{1/2}$ :

$$p_+(a) = 1 - \theta \cdot \beta^{2a^{1/2}} \cdot \delta^a, \quad (8)$$

$$q_{+/(+)}(a) = p_+(a) + \frac{(\theta^{-1} \cdot \beta^{-a^{1/2}} - 1)(1 - p_+(a))^2}{p_+(a)}, \quad (9)$$

where  $\theta = p_0 \cdot b_0 / q_{0/0}^2$ ,  $\beta = q_{0/0} / b_0$  and  $\delta = b_0$  are three constants when a species' distribution is given. This scaling pattern describes species occupancy and local density as a function of the spatial scale,  $a$ , and elucidates the percolation process of how occupied cells joined together when scaling up. That said, species distribution can be heavily manipulated by the spatial scale (resolution) at which it was observed.

For the calculation of species association, we first calculate the probability that both species are absent in the combined  $4a$ -cell  $P_{\bar{A}\bar{B}}(4a)$ . For demonstration, I listed the procedure of calculating this probability in Fig. 1.  $P_{\bar{A}\bar{B}}(4a)$  equals the probability of neither species occurring in a sub-cell  $P_{\bar{A}\bar{B}}(a)$  (Fig. 1(1)), multiplied by the probability of species A being absent in an adjacent sub-cell  $q_{0/0}(a)_A$  (Fig. 1(2)), multiplied by the probability that species B is also absent in this sub-cell given that species A is already known absent  $b'_0(a)_B$  (Fig. 1(3)), multiplied by the probability of species B being absent in an adjacent sub-cell  $q_{0/0}(a)_B$  (Fig. 1(4)), multiplied by the probability that species A is also absent in this adjacent sub-cell given that species B is already known absent  $b''_0(a)_A$

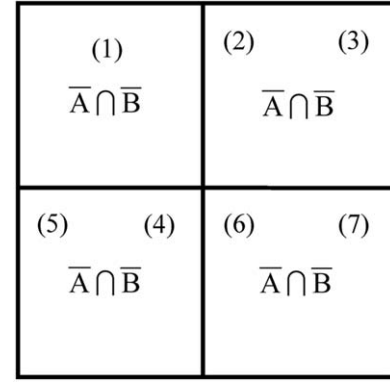


Fig. 1. A schematic illustration of the calculation of the probability  $P_{\bar{A}\bar{B}}(4a)$  that none of the species occurs in the combined  $4a$ -cell. Numbers in brackets indicate the order in the calculation. See text for detail.

(Fig. 1(5)), multiplied by the probability of species A being absent in a chosen sub-cell given that species A is also absent in its two adjacent sub-cells  $b_0(a)_A$  (Fig. 1(6)), and finally multiplied by the probability that species B is absent in this sub-cell given that species B is also absent in its two adjacent sub-cells and also that species A is absent from this sub-cell  $b''_0(a)_B$  (Fig. 1(7)). This procedure yields

$$P_{\bar{A}\bar{B}}(4a) = P_{\bar{A}\bar{B}}(a) \cdot q_{0/0}(a)_A \cdot b'_0(a)_B \cdot q_{0/0}(a)_B \cdot b''_0(a)_A \cdot b_0(a)_A \cdot b''_0(a)_B \quad (10)$$

Probabilities  $q_{0/0}(a)_A$ ,  $q_{0/0}(a)_B$  and  $b_0(a)_A$  are known in Eq. (10) (see Eqs. (2) and (3)). Other three unknown probabilities:  $b'_0(a)_B$ ,  $b''_0(a)_B$  and  $b'_0(a)_A$  can be estimated using the Bayesian rule (note that the formulae for  $b'_0(a)_A$  is symmetric and parallel to the one for  $b'_0(a)_B$ ),

$$b'_0(a)_B = \frac{q_{0/0}(a)_B \cdot p_0(a)_B \cdot P_{\bar{A}\bar{B}}(a)}{q_{0/0}(a)_B \cdot p_0(a)_B \cdot P_{\bar{A}\bar{B}}(a) + q_{0/+}(a)_B \cdot p_+(a)_B \cdot P_{\bar{A}B}(a)}, \quad (11)$$

$$b''_0(a)_B = \frac{q_{0/0}(a)_B^2 \cdot p_0(a)_B \cdot P_{\bar{A}\bar{B}}(a)}{q_{0/0}(a)_B^2 \cdot p_0(a)_B \cdot P_{\bar{A}\bar{B}}(a) + q_{0/+}(a)_B^2 \cdot p_+(a)_B \cdot P_{\bar{A}B}(a)}, \quad (12)$$

where  $P_{\bar{A}\bar{B}}(a)$  and  $P_{\bar{A}B}(a)$  are conditional probabilities defined as follows:

$$P_{\bar{A}\bar{B}}(a) = \frac{P_{\bar{A}\bar{B}}(a)}{P_B(a)} = \frac{1 - p_+(a)_A - p_+(a)_B + R(a)}{1 - p_+(a)_B}, \quad (13)$$

$$P_{\bar{A}B}(a) = \frac{P_{\bar{A}B}(a)}{P_B(a)} = \frac{p_+(a)_B - R(a)}{p_+(a)_B}. \quad (14)$$

The joint occupancy can thus be calculated as

$$P_{A\cap B}(4a) = p_+(4a)_A + p_+(4a)_B - 1 + P_{\bar{A}\bar{B}}(4a). \quad (15)$$

The above equation provides a discrete description of species association. Analogous to Eq. (10), we also have the following probability when combining  $n \times n$  cells together in forming a larger cell:

$$P_{\bar{A}\bar{B}}(n \times n) = P_{\bar{A}\bar{B}}(q_{0/0A} \cdot b'_{0B})^{2(n-1)} (b_{0A} \cdot b''_{0B})^{(n-1)^2}. \quad (16)$$

Let  $n \times n = a$  and we have the continuous version of the scaling pattern of species association,

$$P_{A\cap B}(a) = 1 - \theta_A \beta_A^{2a^{1/2}} \delta_A^a - \theta_B \beta_B^{2a^{1/2}} \delta_B^a + \theta_C \beta_C^{2a^{1/2}} \delta_C^a, \quad (17)$$

where parameters with subscript A and B indicate the parameters as in Eq. (8) for species A and B, respectively; parameters  $\theta_C = P_{\bar{A}\bar{B}} b_{0A} b'_{0B} / (q_{0/0A} b'_{0B})^2$ ,  $\beta_C = q_{0/0A} \cdot b'_{0B} / (b_{0A} \cdot b'_{0B})$  and  $\delta_C = b_{0A} b'_{0B}$  are constants for given distributions of two species. Specifically, if the two species are independent from each other,

we have  $b'_{0B} = q_{0/0B}$  and  $b'_{0B} = b_{0B}$ , and Eq. (17) becomes  $P_{A \cap B}(a) = p_+(a)_A \cdot p_+(a)_B$ , confirming that the definition of species association (as independent, positive or negative) is consistent with Eq. (17).

Therefore, given each species' occupancy and local density, as well as the joint occupancy of two species, we can calculate these three variables at a different scale according to Eqs. (4), (5) and (15), or Eqs. (8), (9) and (17). These equations can thus allow us to extrapolate the pattern of species distributions (occupancy, spatial autocorrelation and species association) across scales. In the following, I show several propositions, as well as results from numerical simulations, of these equations.

### 3. Results

First, I show two propositions regarding two special situations of the above model, which is about the randomness and independence of species distributions.

**Proposition 1.** Random distribution is scale-free and follows a Poisson occupancy–abundance relationship.

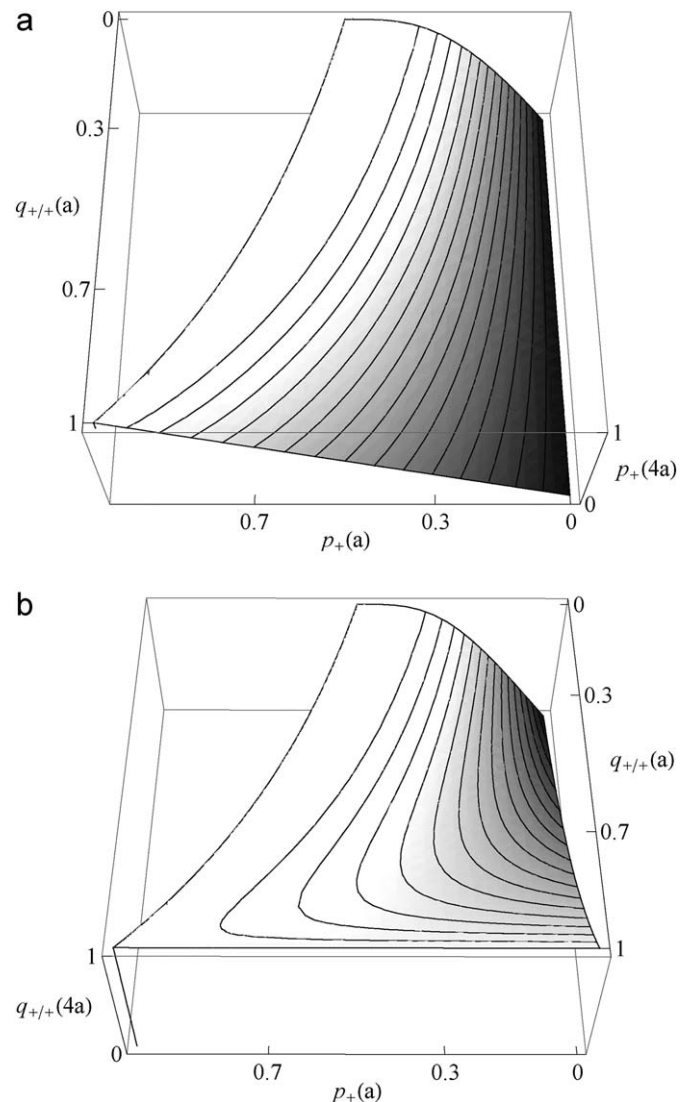
**Proof.** Substituting the condition of a random distribution,  $q_{+/+}(a) = p_+(a)$ , into Eqs. (4) and (5), we have  $q_{+/+}(4a) = p_+(4a)$  and  $1 - p_+(4a) = (1 - p_+(a))^4$ . The former indicates that the species distribution at the scale of  $4a$  still follows a random distribution, i.e. it is scale-free. The latter implies that  $p_0(4a) = p_0(a)^4$ , which infers an exponential function  $p_0(a) = e^{-d \times a}$ , where  $d$  is a positive constant and the negative sign simply indicates the negative relationship between the grain  $a$  and the probability of absence  $p_0(a)$ . For a randomly distributed species, the number of individuals in the cell follows a Poisson distribution (Wright, 1991), from which the probability of presence can be given by  $p_+(a) = 1 - p_0(a) = 1 - e^{-\mu_a}$ . The parameter  $\mu_a$  is the mean number of individuals in the cell (of a size  $a$ ). This complies with the above proposition of an exponential form of the probability of absence. Because  $\mu_a$  equals the population density of the species times the grain, we unveiled that the constant  $d = -(1/a) \ln p_0(a)$  is species density.  $\square$

**Proposition 2.** The independence of two species is scale-free, regardless of the spatial structure (aggregation or segregation) of these two species.

**Proof.** Substituting the condition of  $P_{A \cap B}(a) = P_A(a) \times P_B(a)$  into Eqs. (10)–(14), we have  $P_{A \cap B}(4a) = P_A(4a) \times P_B(4a)$ .  $\square$

These two propositions confirmed our intuitive perception that species random distribution is scale independent and that two species without association (that is, species are independent from each other) is also scale independent. To demonstrate the scaling patterns of species distribution and association from the above Bayesian model, numerical results were presented. First, as shown in Fig. 2, the occupancy and local density at one scale ( $p_+(4a)$  and  $q_{+/+}(4a)$ ) are completely determined by their values at another scale ( $p_+(a)$  and  $q_{+/+}(a)$ ), with no additional parameters involved. Furthermore, both the occupancy  $p_+(a)$  and local density  $q_{+/+}(a)$  limit to one with the increase of grain, indicating that the species distribution converges to randomness at large scales (Fig. 3). However, the category of species distribution, being either aggregated or segregated (separated by the randomness), remains the same across scales (Fig. 3).

For species association, the joint occupancy of species A and B,  $P_{A \cap B}$ , showed a nonlinear relationship when across scales from  $a$  to  $4a$  (Fig. 4a). An interesting result is that even if  $P_{A \cap B}(a) = 0$ , the joint occupancy at larger scales will become positive  $P_{A \cap B}(4a) > 0$ , indicating a higher chance of finding two species co-occurring at

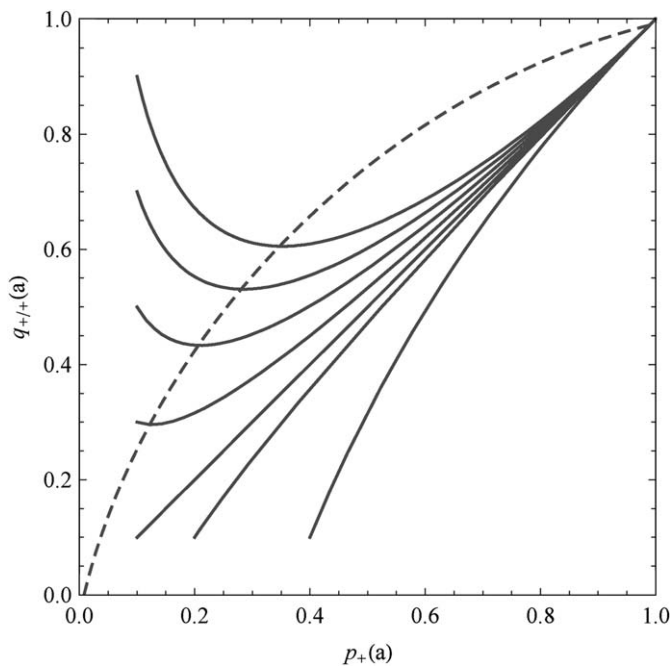


**Fig. 2.** A schematic illustration of (a) the global density (occupancy,  $p_+$ ) and (b) the local density ( $q_{+/+}$ ) by scaling-up from grain  $a$  to  $4a$ , based on Eqs. (4) and (5).

large scales even though they do not overlap at small scales. If we use  $P_{A \cap B}/(P_A \cdot P_B)$  as an index of the intensity of species association ( $P_A$  and  $P_B$  are the occupancies of species A and B, i.e.  $p_+(a)_A$  and  $p_+(a)_B$ ), the category of species association, defined as either positive or negative, remains the same across scales (Fig. 4b). That is, a positive/negative species association will still be positive/negative regardless of the scale of observations. However, a positive association tends to be less positive (i.e. the index approaching to one) when scaling up; similarly, a negative association tends to be less negative when scaling up (Fig. 4b). Therefore, the intensity of species association weakens with the increasing of grain.

The scaling pattern of species distribution and association was shown in Fig. 5. An important result here is that the occupancy  $p_+(a)$ , local density  $q_{+/+}(a)$  and the joint occupancy  $P_{A \cap B}(a)$  all approach one with the scaling-up. However, for aggregated distributions, the local density decreases at first but then increases ultimately at the same rate as  $p_+(a)$  (Fig. 5; see also Fig. 3). The threshold to separate the decreasing part from the increasing part of  $q_{+/+}(a)$  cannot be solved analytically, but an approximate can be obtained by rotating  $\pi/4$  of the ellipse  $(p_+(a) - 1)^2/1.14^2 + q_{+/+}(a)^2/0.89^2 = 1$ , which is insensitive to the initial values of species occupancy and local density (Fig. 3). It





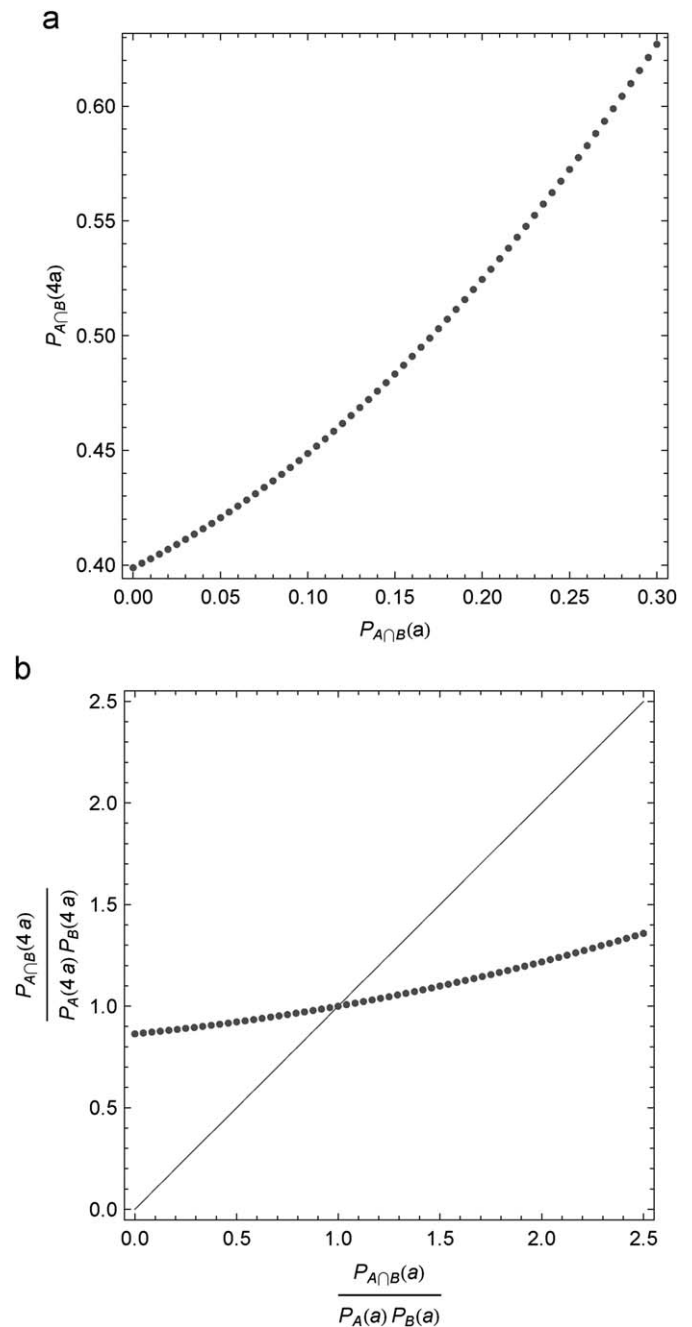
**Fig. 3.** Trajectories (solid lines) of species occupancy  $p_+(a)$  and local density  $q_{+/(a)}$  in a parametric plot, which all lead to the top right with the increase of grain  $a$ . The dashed line indicates the threshold of  $q_{+/(a)}$  where it changes from declining to increasing with the increase of grain.

is worth noting that the local density is not perfectly synchronized with the changing rate of occupancy  $\partial p_+(a)/\partial a$ . However, generally speaking, a lower local density indeed corresponds to a higher changing rate of occupancy, i.e. species distribution becomes strongly scale-dependent when the spatial autocorrelation is weak (or when the local density is low).

#### 4. Discussions

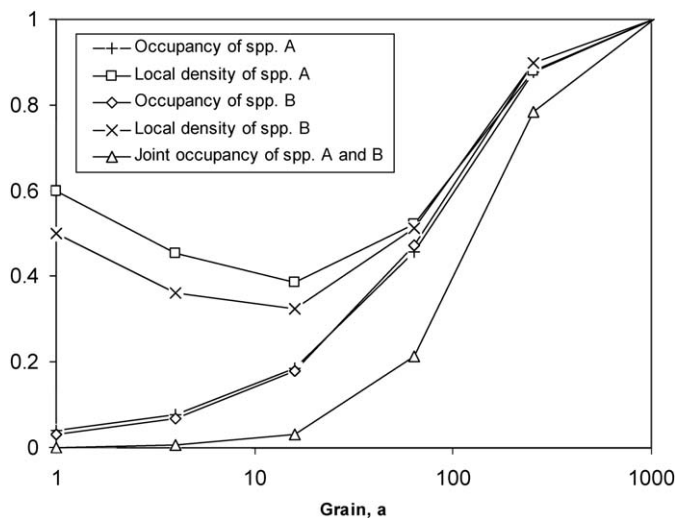
There are currently two types of models available for explaining species occupancy across scales. One is the occupancy–abundance relationship (e.g. Gaston et al., 2006; Hanski and Gyllenberg, 1997; He and Gaston, 2003; Nachman, 1981; Wright, 1991) and is grounded in the ubiquitous positive intraspecific correlation between abundance and range size (Gaston and Blackburn, 2000). Because species abundance should not be a function of the spatial scale, this relationship actually provides a framework of the scaling pattern of species occupancy. The result here showed that Poisson model, the simplest occupancy–abundance relationship (Wright, 1991), is a special case (randomness) of current model, suggesting that the model here is consistent with those models based on the occupancy–abundance relationship. The other type of model is based on the scaling pattern of occupancy (Hartley and Kunin, 2003; Hui et al., 2006, 2009; Kunin, 1998), describing how adjacent occupied cells merge with increasing grain. The Bayesian model presented here is a generalized model of the scaling pattern of occupancy (Hui et al., 2006). Therefore, random distribution described by current model can be served as a bridge between these two types of models.

Empirical occupancy data are commonly logistically related to the log-transformed grain (Gatrell et al., 1996; He and Gaston, 2000; Hui and McGeoch, 2007b; Perry, 1995; Fig. 5). This logistical shape (or “S” shape) describes a rapid phase transition from a highly fragmented distribution to a highly connected one when scaling up, and reflects the percolation process of the scaling



**Fig. 4.** (a) The joint occupancy of species A and B at the scale of  $4a$ ,  $P_{A \cap B}(4a)$ , as a function of the joint occupancy at the scale of  $a$ ,  $P_{A \cap B}(a)$ , with  $p_+(a)_A = 0.4$ ,  $q_{+/(a)}_A = 0.6$ ,  $p_+(a)_B = 0.3$  and  $q_{+/(a)}_B = 0.5$ . (b) A comparison of the intensity of species association  $P_{A \cap B}/(P_A P_B)$  with  $y = x$  (straight line) when scaling-up from  $a$  to  $4a$ .

pattern of species distribution (He and Hubbell, 2003; Hui and McGeoch, 2007b; Plotkin et al., 2002). Percolation process is a mathematical description of the random spread of fluid through a medium (Sahimi, 1994; Smythe and Wierman, 1978), and has been shown overwhelming in the processes of pattern formation in spatial ecology (Tilman et al., 1997). With an increase in grain, adjacent occupied cells merge as a cluster, which is analogous to the continuum and band percolation (Plotkin et al., 2002), whereas with a decrease in grain, the percolation process is essentially similar to the construction of a random Cantor set (a fractal percolation; Mandelbrot, 1983). The model here complies with the percolation process in understanding species



**Fig. 5.** An example of the scaling patterns of species occupancy ( $p_{+}(a)_A$  and  $p_{+}(a)_B$ ), autocorrelation (local densities  $q_{+/+}(a)_A$  and  $q_{+/+}(a)_B$ ) and species association (joint occupancy  $P_{A \cap B}(a)$ ).

spatial patterns. Furthermore, another commonly observed scaling pattern of occupancy is a log–log linear relationship between grain and species occupancy, known as the area-of-occupancy (Gaston and Fuller, 2009; Hartley and Kunin, 2003; Kunin, 1998), which reflects the fractal nature of species distribution. Even though the current model does not infer a strict log–log linear scaling pattern of occupancy, a log–log transformation can indeed largely linearize the model prediction. This suggests that the log–log linear relationship might not be the full-scale scaling pattern of occupancy. Other studies indeed show that the fractal nature of species distribution breaks down at small scales (Hui and McGeoch, 2007c).

The scale-dependence of species aggregation and spatial autocorrelation has been well recognized (Dungan et al., 2002). This scale-dependent pattern of aggregation is part of a well-known modifiable areal unit problem (MAUP; Openshaw, 1984), describing the variation in the patterns of species distributions as a result of “modifiable” sampling scales (Dungan et al., 2002; Jelinski and Wu, 1996). Following Fotheringham’s (1989) call of shifting focus from seeking scale-invariant measure to quantifying rates of change in species distributional measures (see also Wiegand and Moloney, 2004), the model presented here revealed the rate of change in species local density (join-count statistics) across scales. First, the scales of the lowest local density in general correspond with the inflection scales of the occupancy (i.e. where the rate of change of occupancy is highest) (Fig. 5; see also He and Hubbell, 2003; Hui et al., 2006). This is because, at the same scale, species with a lower local density is more widespread than species with a similar occupancy but a higher local density. As a result, the species with a lower local density will have a higher occupancy when scaling up. That is, species’ range size (measured by occupancy) becomes highly scale-dependent when the spatial autocorrelation is weak. Second, the scales of the lowest local density can be depicted by the threshold of percolation (dashed line in Fig. 3). Specifically, the occupied cells of species with low occupancy but high local density first are quickly clustered together when scaling up, causing a decline of local density due to the exhaustion of neighbouring occupied cells. The occupancy then climbs up when the clusters start forming super clusters, and therefore a threshold of the minimum local densities is formed (dashed line in Fig. 3). Moreover, the intensity of species non-randomness ( $q_{+/+}/p_{+}$ ) approaches one (randomness) when scaling up. It implies that this pattern variance (spatial

autocorrelation), which results from intraspecific interactions, is often distance-limited (Dieckmann et al., 2000), and therefore becomes difficult to detect at large scales.

Species association is an important concept for understanding patterns in ecological communities (e.g., co-occurrence and nestedness; Bell, 2005; Gotelli and Graves, 1996; Ulrich and Gotelli, 2007). It also further helps to identify the interspecific interactions (e.g. competition or facilitation) behind community patterns other than formed by chance alone (Bell, 2005; Connor and Simberloff, 1979; Diamond, 1975; Sfenthourakis et al., 2005). Results here suggested that broad-scale samples could be difficult to identify community patterns caused mainly by distance-limited interspecific interactions. Heard and Buchanan (1998) indeed found that although no evidence for interactions among nipple gall-makers on different leaves, significant interactions can be detected at a smaller spatial scale. The effect of spatial scales on two- and multi-species interactions and assembly patterns warrants further investigations.

In conclusion, the intensity of species spatial correlation and association declines with the increase of grain, i.e. from fine (small) to coarse (large) scales, even though the category of species distribution (aggregated or segregated) and association (positive or negative) remains the same. Random distribution and species independence are scale-free patterns. When scaling up, species distributions become random, and species associations between two species become independent even though they have an increasing joint occupancy. By describing spatial patterns in terms of occupancy, first-order spatial autocorrelation, and species association, the Bayesian model can grasp the statistical essence of species scaling patterns and thus represents a step towards a general scaling model for species distributional structure and assembly in ecological communities.

## Acknowledgements

I am grateful to M.A. McGeoch, B. Laniewski, K.J. Gaston, A.L. Šizling, M.A. Lewis and two anonymous referees for constructive comments, the NRF Blue Skies Research Programme and the DST-NRF Centre of Excellence for Invasion Biology for financial support.

## References

- Anselin, L., 1995. Local indicators of spatial association—LISA. *Geogr. Anal.* 27, 93–116.
- Bell, G., 2005. The co-distribution of species in relation to the neutral theory of community ecology. *Ecology* 86, 1757–1770.
- Connor, E.F., Simberloff, D., 1979. The assembly of species communities: change or competition. *Ecology* 60, 1132–1140.
- Diamond, J.M., 1975. Assembly of species communities. In: Cody, M.L., Diamond, J.M. (Eds.), *Ecology and Evolution of Communities*. Harvard University Press, Cambridge, pp. 342–444.
- Dieckmann, U., Law, R., Metz, A.J.A., 2000. *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*. Cambridge University Press, Cambridge.
- Dungan, J.L., Perry, J.N., Dale, M.R.T., Legendre, P., Citron-Pousty, S., Fortin, M.-J., Jakomulska, A., Miriri, M., Rosenberg, M.S., 2002. A balanced view of scale in spatial statistical analysis. *Ecography* 25, 626–640.
- Ellner, S.P., 2001. Pair approximation for lattice models with multiple interaction scales. *J. Theor. Biol.* 210, 435–447.
- Fahrig, L., Nutton, W.K., 2005. Population ecology in spatial heterogeneous environments. In: Lovett, G.M., Jones, C.G., Turner, M.G., Weathers, K.C. (Eds.), *Ecosystem Function in Heterogeneous Landscapes*. Springer, Berlin, pp. 95–118.
- Fortin, M.-J., Dale, M.R.T., 2005. *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press, Cambridge.
- Fortin, M.-J., Dale, M.R.T., van Hoef, J., 2002. Spatial analysis in ecology. In: El-Shaarawi, A.H., Piegorisch, W.W. (Eds.), *Encyclopedia of Environmetrics*. Wiley, New York, pp. 2051–2058.
- Fotheringham, A.S., 1989. Scale-independent spatial analysis. In: Goodchild, M.F., Gopal, S. (Eds.), *Accuracy of Spatial Databases*. Taylor & Francis, London, pp. 221–228.

- Gaston, K.J., Blackburn, T.M., 2000. Patterns and Process in Macroecology. Blackwell Science, Oxford.
- Gaston, K.J., Fuller, R.A., 2009. The sizes of species' geographic ranges. *J. Appl. Ecol.* 46, 1–9.
- Gaston, K.J., Borges, P.V., He, F., Gaspar, C., 2006. Abundance, spatial variance and occupancy: arthropod species distribution in the Azores. *J. Anim. Ecol.* 75, 646–656.
- Gatrell, A.C., Bailey, T.C., Diggle, P.J., Rowlingson, B.S., 1996. Spatial point pattern analysis and its application in geographical epidemiology. *Trans. Inst. Br. Geogr.* 21, 256–274.
- Gotelli, N.J., Graves, G.R., 1996. Null Models in Ecology. Smithsonian Institution Press, London.
- Hanski, I., Gyllenberg, M., 1997. Uniting two general patterns in the distribution of species. *Science* 284, 397–400.
- Harte, J., Conlisk, E., Ostling, A., Green, J.L., Smith, A.B., 2005. A theory of spatial structure in ecological communities at multiple spatial scales. *Ecol. Monogr.* 75, 179–197.
- Hartley, S., Kunin, W.E., 2003. Scale dependency of rarity, extinction risk, and conservation priority. *Conserv. Biol.* 17, 1559–1570.
- He, F., Gaston, K.J., 2000. Occupancy–abundance relationships and sampling scales. *Ecography* 23, 503–511.
- He, F., Gaston, K.J., 2003. Occupancy, spatial variance, and the abundance of species. *Am. Nat.* 162, 366–375.
- He, F., Hubbell, S.P., 2003. Percolation theory for the distribution and abundance of species. *Phys. Rev. Lett.* 91, 198103.
- Heard, S.B., Buchanan, C.K., 1998. Larval performance and association within and between two species of hackberry nipple gall insects, *Pachypsylla* spp. (Homoptera: psyllidae). *Am. Midl. Nat.* 140, 351–357.
- Holt, A.R., Gaston, K.J., He, F., 2002. Occupancy–abundance relationships and spatial distribution: a review. *Basic Appl. Ecol.* 3, 1–13.
- Hui, C., 2008. Crossing the borders of spatial analysis and modelling: a rethink. In: Kelly, J.T. (Ed.), *Progress in Mathematical Biology Research*. Nova Science, Hauppauge, NY, pp. 191–218.
- Hui, C., Li, Z., 2004. Distribution patterns of metapopulation determined by Allee effects. *Popul. Ecol.* 46, 55–63.
- Hui, C., McGeoch, M.A., 2007a. A self-similarity model for the occupancy frequency distribution. *Theor. Popul. Biol.* 71, 61–70.
- Hui, C., McGeoch, M.A., 2007b. Capturing the “droopy-tail” in the occupancy–abundance relationship. *Écoscience* 14, 103–108.
- Hui, C., McGeoch, M.A., 2007c. Modeling species distributions by breaking the assumption of self-similarity. *Oikos* 116, 2097–2107.
- Hui, C., McGeoch, M.A., 2008. Does the self-similarity species distribution model lead to unrealistic predictions? *Ecology* 89, 2946–2952.
- Hui, C., McGeoch, M.A., Warren, M., 2006. A spatially explicit approach to estimating species occupancy and spatial correlation. *J. Anim. Ecol.* 75, 140–147.
- Hui, C., McGeoch, M.A., Reyers, B., le Roux, P.C., Greve, M., Chown, S.L., 2009. Extrapolating population size from the occupancy–abundance relationship and the scaling pattern of occupancy. *Ecol. Appl.*, in press.
- Hurlbert, S.H., 1990. Spatial distribution of the montane unicorn. *Oikos* 58, 257–271.
- Ives, A.R., 1991. Aggregation and coexistence in a Carrion fly community. *Ecol. Monogr.* 61, 75–94.
- Iwasa, Y., Sato, K., Nakashima, S., 1991. Dynamic modeling of wave regeneration (Shimagare) in subalpine Abies forests. *J. Theor. Biol.* 152, 143–158.
- Jelinski, D.E., Wu, J., 1996. The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecol.* 11, 129–140.
- Kunin, W.E., 1998. Extrapolating species abundance across spatial scales. *Science* 281, 1513–1515.
- Levin, S.A., 1992. The problem of pattern and scale in ecology. *Ecology* 73, 1943–1967.
- Mandelbrot, B., 1983. *The Fractal Geometry of Nature*. Freeman, San Francisco.
- Matsuda, H., Ogita, A., Sasaki, A., Sato, K., 1992. Statistical mechanics of population: the lattice Lotka–Volterra model. *Prog. Theor. Phys.* 88, 1035–1045.
- McGeoch, M.A., Gaston, K.J., 2002. Occupancy frequency distributions: patterns, artefacts and mechanisms. *Biol. Rev.* 77, 311–331.
- Molles, M.C., 1999. *Ecology: Concepts and Application*. McGraw-Hill, New York.
- Nachman, G., 1981. A mathematical model of the functional relationship between density and spatial distribution of a population. *J. Anim. Ecol.* 50, 453–463.
- Openshaw, S., 1984. *The Modifiable Areal Unit Problem*. GeoBooks, Norwick.
- Pacala, S.W., Levin, S.A., 1997. Biological generated spatial pattern and the coexistence of competing species. In: Tilman, D., Kareiva, P. (Eds.), *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*. Princeton University Press, Princeton, pp. 204–232.
- Perry, J.N., 1995. Spatial analysis by distance indices. *J. Anim. Ecol.* 64, 303–314.
- Perry, J.N., Liebold, A.M., Rosenberg, M.S., Dungan, J., Miriti, M., Jakomulska, A., Citron-Pousty, S., 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. *Ecography* 25, 578–600.
- Plotkin, J.B., Chave, J., Ashton, P.S., 2002. Cluster analysis of spatial patterns in Malaysian tree species. *Am. Nat.* 160, 629–644.
- Sahimi, M., 1994. *Applications of Percolation Theory*. Taylor & Francis, London.
- Sato, K., Iwasa, Y., 2000. Pair approximation for lattice-based ecological models. In: Dieckmann, U., Law, R., Metz, J.A.J. (Eds.), *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*. Cambridge University Press, Cambridge, pp. 341–359.
- Sfenthourakis, S., Tzanatos, E., Giokas, S., 2005. Species co-occurrence: the case of congeneric species and a causal approach to patterns of species association. *Global Ecol. Biogeogr.* 15, 39–49.
- Shorrocks, B., Sevenster, J.G., 1995. Explaining local species-diversity. *Proc. R. Soc. London B* 260, 305–309.
- Smythe, R.T., Wieman, J.C., 1978. *First-passage Percolation on the Square Lattice*. Springer, Berlin.
- Stone, L., Roberts, A., 1990. The checkerboard score and species distributions. *Oecologia* 85, 74–79.
- Storch, D., Šizling, A.L., Reif, J., Polechová, J., Šizlingová, E., Gaston, K.J., 2008. The quest for a null model for macroecological patterns: geometry of species distributions at multiple spatial scales. *Ecol. Lett.* 11, 771–784.
- Tilman, D., Lehman, C.L., Kareiva, P., 1997. *Population dynamics in spatial habitats*. In: Tilman, D., Kareiva, P. (Eds.), *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*. Princeton University Press, Princeton, pp. 3–20.
- Ulrich, W., Gotelli, N.J., 2007. Disentangling community patterns of nestedness and species co-occurrence. *Oikos* 116, 2053–2061.
- Wertheim, B., Sevenster, J.G., Eijs, I.E.M., van Alphen, J.J.M., 2000. Species diversity in a mycophagous insect community: the case of spatial aggregation vs. resource partitioning. *J. Anim. Ecol.* 69, 335–351.
- Wiens, J.A., 2000. Ecological heterogeneity: ontogeny of concepts and approaches. In: Hutchings, M.J., Jones, E.A., Stewart, A.J.A. (Eds.), *The Ecological Consequences of Environmental Heterogeneity*. Blackwell Science, Oxford, pp. 9–31.
- Wiegand, T., Moloney, K.A., 2004. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos* 104, 209–229.
- Wilson, R.J., Thomas, C.D., Fox, R., Roy, D.B., Kunin, W.E., 2004. Spatial patterns in species distributions reveal biodiversity change. *Nature* 432, 393–396.
- Wright, D.H., 1991. Correlations between incidence and abundance are expected by chance. *J. Biogeogr.* 1, 463–466.
- Wu, J., 2004. Effects of changing scale on landscape pattern analysis: scaling relations. *Landscape Ecol.* 19, 125–138.